ECE 506: Homework #1

Problem #1. Basics of the use of the l_1 norm.
1(a) Consider the following constraint on the l_1 norm:

$$\sum_{i=1}^{i=n} |x_i| \leq 1. \tag{1}$$

Draw the feasible region that satisfies the constraint for $n = 2$.

1(b) Restate (1) in standard form using: $c_i(x)$, $i = 1, 2, \ldots, 4$.
1(c) Consider the general case where

$$\sum_{i=1}^{i=n} |x_i| \leq M \tag{2}$$

where $M < n$ and i is an integer.

1. List the corner points.
2. For each corner point, show the count of the non-zero entries.

Hint: Corner points satisfy $x_i = 0$ for some indices. For example, in 2D, we have that $x_1 = 0, x_2 = 1$ is a corner point with one non-zero entry.

1(d) Restate (2) using $c_i(x)$ for $n = 3$. How many $c_i(x)$ constraints do you need for arbitrary n? Why do you think that the use of the l_1-norm requires the use of approximations of the constraints?

Problem #2. Unconstrained optimization.
Consider the following ideal, convex function:

$$f(x_1, x_2) = (x_1 - 1)^2 + 10 \cdot (x_2 - 2)^2 \tag{3}$$

1. Plot the contours of the function and its gradient.
2. Compute the point where $\nabla f(x) = 0$.
3. Solve the unconstrained optimization problem:

$$\min_x \quad f(x_1, x_2). \tag{4}$$

4. Verify $\nabla f(x) = 0$ at the optimal point.

Problem #3. Constrained optimization.
Consider the following ideal, convex optimization problem:

$$\min_{x} \quad f(x_1, x_2) \tag{5}$$

subject to:

$$\sum_{i=1}^{i=n} |x_i| \leq 1.$$

3(a) Let $f(x_1, x_2) = (x_1 - 4)^2 + (x_2 - 4)^2$.

1. Plot the contours of the function and its constraints.
2. Solve the unconstrained optimization problem:

$$\min_{x} \quad f(x_1, x_2). \tag{6}$$

3. Solve the constraint optimization problem by finding the point in the line constraint that is closest to the unconstrained optimal point.

3(b) Let $f(x_1, x_2) = (x_1 - 10)^2 + x_2^2$.

1. Plot the contours of the function and its constraints.
2. Solve the unconstrained optimization problem:

$$\min_{x} \quad f(x_1, x_2). \tag{7}$$

3. Solve the constraint optimization problem by finding the point that is closest to the unconstrained optimal point.

3(c) Let $f(x_1, x_2) = (x_1 - 0.5)^2 + x_2^2$.

1. Plot the contours of the function and its constraints.
2. Solve the unconstrained optimization problem:

$$\min_{x} \quad f(x_1, x_2). \tag{8}$$

3. Based on your contour-plot show that the optimal point cannot be on the boundary.

3(d) Vertify that the KKT conditions are satisfied at the optimal points for 3(a), 3(b), and 3(c).

KKT Summary: In constrained-optimization, the KKT conditions are always satisfied at the optimal point. Note that this is not enough to determine if a point is optimal. If the KKT conditions are satisfied, we cannot infer that we have an optimal point. To understand the KKT conditions, we need to apply them. Here is an intuitive summary of how to apply and verify the KKT conditions:

```
IF (the solution is at an interior point)
   THEN The gradient of f should be zero at that point.

IF (the solution is on a boundary point) AND (LICQ holds there)
  THEN
     The Lagrangian condition is satisfied at that point.
```

To understand the KKT conditions, note that they only tell you when you can reject a candidate optimal point. To see how to do this, let us review the elements of a contra-positive proof. Suppose that we have:

```
IF A THEN B
```

Then the statement is equivalent to:

```
IF (NOT B) THEN (NOT A)
```

In other words, since A implies B, if B does not hold, it must be that A does not hold either. If we apply this statement to the first statement, we have that if the gradient of f is not zero in the interior, we can conclude that the optimal solution is not in the interior. If we apply this statement to the second statement, we then have that if the Lagrangian condition is not satisfied, we then have that either the solution is not on the boundary or the LICQ does not hold.

If the KKT conditions are satisfied, then we MAY be at an optimal point. There are no guarantees that we will be optimal. Back in our example, if A implies B, we cannot say that B implies A.

Next, we explain each condition carefully. For the first condition, if x^* is inside the feasible region, we require that $\nabla f(x^*) = 0$. Again, this is not enough to guarantee that we are at an optimal point. However, if this condition is violated, we are clearly not at an optimal point in the interior.

If the solution is not inside the feasible region, then it could be on the boundary of the feasible region. In this case, we need to first check the LICQ condition before we look at the full KKT conditions.

For the LICQ condition, consider the vectors evaluated at the candidate optimal point. Here, we are only concerned with the constraints that are **active**. We use the term **active** to describe the constraints that satisfy $c_i(x^*) = 0$. Thus, if a solution is on a line, the $c_i(x)$ that gives the equation of the line will satisfy $c_i(x^*) = 0$. For a corner, we will have the $c_i(x^*) = c_j(x^*) = 0$ where c_i and c_j represent each intersecting line. The LICQ requires that $\nabla c_i(x^*)$ from the active constraints remain **linearly independent**. To prove LICQ, you need to show that

$$\sum_{\textbf{Active } i} a_i \nabla c_i(x^*) = 0 \tag{9}$$

is only satisfied if $a_i = 0$ for all i. Intuitively, we say that the **active** constraint gradients are independent from each other. In other words, the constraints pull you into different directions inside the boundary.

To check the Lagrangian condition, we form the Lagrangian:

$$\mathcal{L}(x, \lambda) = f - \sum_{\textbf{Active } i} \lambda_i c_i(x). \tag{10}$$

The Lagrangian condition requires that we can solve

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \tag{11}$$

at the optimal point x^* for some $\lambda_i^* \geq 0$.

It is important to understand what (11) is saying. Basically, it is saying that going inside the feasible region will only increase the function. Note that ∇f is the direction where the function is increasing. Furthermore, the solution implies that

$$\nabla f(x^*) - \sum_{\textbf{Active } i} \lambda_i^* \nabla c_i(x^*) = 0. \tag{12}$$

In turn, this implies that:

$$-\nabla f(x^*) = -\sum_{\textbf{Active } i} \lambda_i^* \nabla c_i(x^*). \tag{13}$$

Thus, the descent direction of $-\nabla f(x^*)$ will move us outside the feasible region.

Contour and gradient plots in Python

For information on how to plot contours, the following links can help a lot:

- A simple unofficial tutorial on how to generate contours in Python is given at `https://www.python-course.eu/matplotlib_contour_plot.php`.
- The official tutorial of how to plot 2D contours from regularly-spaced points.
- The official tutorial of how to plot 2D contours from irregularly-spaced points (needed later).
- An official advanced tutorial for using advanced options for contour plots can be found at advanced contour tutorial.
- An official simple demo of how to use `quiver` to plot gradient fields.
- In Matplotlib, you can plot the contours followed by the gradient and they will appear together. In other words, this is equivalent to having `hold on` as the default behavior.

Advanced Demo for Optimization Methods in Python

You can find a very nice demonstration of how to produce convergence videos of several unconstrained optimization algorithms.